

CLAIMS

- 1) A method for finding translation initiation codons in a nucleotide sequence, comprising:
 - a) analyzing a first data set to measure a combination of features of initiator codons and pseudoinitiator codons and to produce a set of numerical values for said combination of features; and
 - b) evaluating scoring functions by reading a sequence in the vicinity of an ATG triplet and using said scoring functions and said scoring function's parameters to return a numerical score that quantifies how much said ATG triplet resembles an initiator codon; and
 - c) generating a quadratic discriminant function through selection of a combination of feature variables that optimally classifies ATG triplets in a nucleotide sequence as initiator codons or as pseudoinitiator codons based on the output of said scoring functions and through the use of Quadratic Discriminant Analysis; and
 - d) using said quadratic discriminant function to analyze a second data set of nucleotide sequences by evaluating at least one scoring function for each ATG triplet in said sequences and to calculate the probability of an initiator codon at a position using the output of said analysis.
- 2) A method for finding translation initiation codons in a nucleotide sequence, as recited in claim 1, wherein said combination of features from step a) comprises at least two of the features provided in Table 1.

- 3) A method for finding translation initiation codons in a nucleotide sequence, as recited in claim 1, wherein said scoring functions from step d) comprise at least two of the scoring functions provided in Table 2.
- 4) A method for finding translation initiation codons in a nucleotide sequence, as recited in claim 1, wherein said combination of feature variables from step c) comprises any combination of at least two of the feature variables provided in Table 3 wherein the combination comprises one feature variable from each of any two feature variable classes and results in a correlation coefficient for the feature variable combination of greater than 0.9.
- 5) A method for finding translation initiation codons in a nucleotide sequence, as recited in claim 1, wherein said combination of feature variables from step c) comprises any combination of at least two of the feature variables provided in Table 3 wherein the combination comprises one feature variable from each of any two feature variable classes and results in a correlation coefficient for the feature variable combination of greater than 0.8.